LINKÖPINGS UNIVERSITET

Matematiska institutionen

17 augusti 2020, kl. 08.00-12.00
Examinator: Jörg-Uwe Löbus

Alla hjälpmedel är tillåtna.

Betygsgränser 3: minst 8 poäng, 4: minst 11.5 poäng, 5: minst 15 poäng (av 18)

(1) Låt $X_1, X_2, \ldots, X_{100}$ vara oberoende stokastiska variabler, alla med samma fördelning given av $P(X_i = -1) = P(X_i = 0) = P(X_i = 1) = 1/3$.

(a) Bestäm väntevärdet och variansen för X_1. (1.5p)

(b) Låt $Y = X_1 + \ldots + X_{100}$. Bestäm approximativt $P(Y > 10)$. (1.5p)

Ledning: Använd centrala gränsvärdessatsen.

(2) Ett företag som tillverkar batterier av en viss typ har tillverkningen förlagd till tre olika fabriker. Fabrik A står för 50 % av tillverkningen, fabrik B för 20 % och fabrik C för 30 %. Man vet att ett batteri från fabrik A har sannolikheten 95 % att räcka mer än 10 drifttimmar. Motsvarande sannolikheter för fabrikerna B och C är 97 % resp. 98 %. Man har blandat batterier från de tre fabrikerna i ett stort centralt lager.

(a) Vad är sannolikheten att ett batteri som tas på måfå ur lagret ska räcka mer än 10 drifttimmar? (1p)

(b) Man tar på måfå ett batteri ur lagret och finner att det räcker mer än 10 drifttimmar. Vad är sannolikheten för att det tillverkats i fabrik A? (1p)

(c) Man tar på måfå ett batteri ur lagret och finner att det räcker mindre än 10 drifttimmar. Vad är sannolikheten för att det tillverkats i fabrik A? (1p)

(3) Två personer A och B har stämt möte på ett kafé "litet efter klockan åtta". A anländer X minuter efter åtta och B anländer Y minuter efter åtta. Anta nu att X är $U(0,5)$-fördelad och Y är $U(0,7)$-fördelad (enhet är minuter). X och Y antas vara oberoende.

(a) Bestäm den simultana täthetsfunktionen av (X, Y). (1p)

(b) Vad är sannolikheten att A får vänta på B, dvs. vad är $P(X < Y)$? (2p)

(4) En hästhopperska vill att hennes ryttarprogram skall ta 30 sekunder. Hon utför programmet 5 gånger med sluttiderna (enhet: sek)

$$29.8 \quad 35.1 \quad 31.9 \quad 27.4 \quad 32.5\ .$$

Antag att värden är oberoende och varierar enligt normalfördelning.

(a) Ange det generella (tvåsidiga) konfidensintervallet I_μ för väntevärdet i en normalfördelad population där standardavvikelsen är okänd. (1p)

(b) Beräkna ett tvåsidigt 95% konfidensintervall för den förväntade tiden som programmet tar. (1.5p)

(c) Får man hävda att det är statistiskt säkerställt att programmets förväntad tid är kortare än 34 sekunder, på konfidensnivå 95%? (0.5p)

(5) De två sämsta segertiderna på Stockholm Marathon har registrerats under de varmaste dagarna (1982 och 2007). I nedanstående tabell finns temperatur och segerrtid (i minuter) till och med 2008.

År	Temp	Tid	År	Temp	Tid
1979	19	137.58	1994	17	134.10
1980	17	135.82	1995	17	134.48
1981	19	133.43	1996	17	135.08
1982	31	139.33	1997	26	137.37
1983	15	131.62	1998	19	136.20
1984	19	133.78	1999	23	134.87
1985	22	137.30	2000	16	138.82
1986	11	132.55	2001	17	138.28
1987	7	133.87	2002	25	138.33
1988	11	134.43	2003	16	138.23
1989	13	133.57	2004	18	136.20
1990	19	133.07	2005	13	133.50
1991	12	132.63	2006	22	137.02
1992	28	135.97	2007	29	140.93
1993	22	136.50	2008	25	136.05

Låt x_i beteckna temperatur och y_i segertid och anta att data följer modellen enkel linjär regression (Stockholms universitet)

$$Y_i = \alpha + \beta x_i + \varepsilon_i \,, \quad i = 1, 2, \ldots, n \,,$$

där $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ är oberoende och normalfördelade med väntevärde 0 och varians σ^2.

(a) Skatta standardavvikelsen σ. (1p)

(b) Bestäm ett tvåsidigt 95%-konfidensintervall för koefficienten β. (1p)

(c) Antag att temperaturen under Stockholm Marathon nästa år är $x_0 = 10\ {}^oC$. Ange ett 95%-prediktionsintervall för segertiden Y_0 vid temperaturen x_0. (1p)

Räknehjälp: $\sum x_i = 565$, $\sum y_i = 4070.91$, $\sum x_i^2 = 11577$, $\sum y_i^2 = 552563.70$, $\sum x_i y_i = 76909.63$, $\alpha^*_{obs} = 130.85$, $\beta^*_{obs} = 0.257$. Observera att α^*_{obs} kan ha ett avrundningsfel.

(6) Låt X vara en stokastisk variabel med täthetsfunktion

$$f_X(x) = \begin{cases} \dfrac{x}{\theta^2} e^{-x/\theta} & \text{för} \quad x \geq 0 \\ 0 & \text{för} \quad x < 0 \end{cases} .$$

(a) Härled maximumlikelihood-skattningen av θ baserad på observationer $x_1, \ldots, x_n$ av X. (1.5p)

(b) Är skattningen i (a)-delen väntevärdesriktig? (1.5p)

Lösningar

(1) Vi har en följd av oberoende och likafördelade stokastiska variabler $X_1, \ldots, X_{100}$ med samma sannolikhetsfunktion $P(X_i = -1) = P(X_i = 0) = P(X_i = 1) = 1/3$.

(a) $\mu = E[X_1] = 0$, $\sigma^2 = Var(X_1) = 2/3$

(b) Här ska vi sätta $Y = X_1 + \ldots + X_{100}$ och approximera $P(Y > 10)$, och då behöver vi använda centrala gränsvärdessatsen.

$$\begin{aligned} P(Y > 10) &= P\left(\sum_{i=1}^{100} X_i > 10\right) = P\left(\frac{\sum_{i=1}^{100} X_i - 100\mu}{\sigma\sqrt{100}} > \frac{10 - 100\mu}{\sigma\sqrt{100}}\right) \\ &\approx P(Z > 1.22) = 1 - \Phi(1.22) = 1 - 0.8888 = 0.1112 \end{aligned}$$

där $Z \sim N(0,1)$.

(2) Låt A, B och C beteckna händelserna att ett på måfå valt batteri kommer från fabrik A, B eller C. Vidare, låt R beteckna händelsen att valt batteri räcker mer än 10 drifttimmar.

(a) Total sannolikhet:

$$\begin{aligned} P(R) &= P(R|A) \cdot P(A) + P(R|B) \cdot P(B) + P(R|C) \cdot P(C) \\ &= 0.95 \cdot 0.5 + 0.97 \cdot 0.2 + 0.98 \cdot 0.3 = 0.963\,. \end{aligned}$$

(b) Definition betingad sannolikhet:

$$P(A|R) = \frac{P(A \cap R)}{P(R)} = \frac{0.475}{0.963} = 0.49325\,.$$

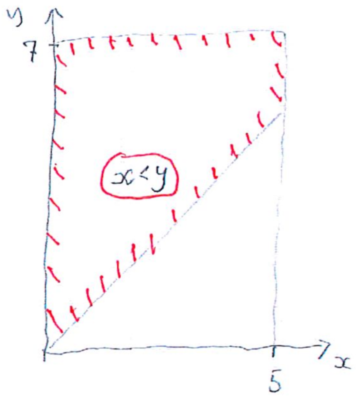
(c) Bayes' formel:

$$P(A|R^c) = \frac{P(R^c|A) \cdot P(A)}{P(R^c)} = \frac{P(R^c|A) \cdot P(A)}{1 - P(R)} = \frac{0.05 \cdot 0.5}{0.037} = 0.6757\,.$$

(3) (a) Oberoendet av X och Y medför att

$$f(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} \frac{1}{5} \cdot \frac{1}{7} & \text{om} \quad 0 \le x \le 5, \quad 0 \le y \le 7 \\ 0 & \text{annars} \end{cases}\,.$$

(b) Sökt är $P(X < Y)$.



Vi får

$$P(X<Y)=\frac{\text{markerad area}}{\text{total area}}=\frac{35-5\cdot 5/2}{35}=\frac{22.5}{35}=\frac{9}{14}\,.$$

(4) (a) Konfidensintervallet för μ är

$$I_\mu=\left(\bar{x}-t(n-1)_{\alpha/2}\cdot\frac{s}{\sqrt{n}}\,,\ \bar{x}+t(n-1)_{\alpha/2}\cdot\frac{s}{\sqrt{n}}\right)\,.$$

(b) Låt x beteckna tiden för ryttarprogrammet i sekunder. Från data beräknar vi på sedvanligt sett att $\bar{x}=31.34$ och $s=2.90$. Vi har ett litet stickprov med okänd standardavvikelse. Under antagande om normalfördelning kan vi därför använda oss av t-fördelningen.

Vi har att $n=5$, $\alpha=0.05$ och $t(n-1)_{\alpha/2}=2.776$. Med detta insatt ovan får vi ett 95% tvåsidigt konfidensintervall:

$$I_\mu=(27.74,34.94)\,.$$

(c) Eftersom $34\in I_\mu$, så är det med konfidensgrad 95% inte statistiskt säkerställt att programmets förväntad tid är kortare än 34 sekunder.

(5) (a) Enligt formeln är skattningen $s=\sigma^*_{obs}=1.808$.

(b) Vi har $S_{xx}=\sum x_i^2-n\cdot\bar{x}^2=11577-30\cdot(565/30)^2=936.17$

$$\begin{aligned}I_\beta&=\beta^*_{obs}\pm t_{a/2}(n-2)\cdot\frac{s}{\sqrt{S_{xx}}}\\&=\left(0.257-2.05\cdot\frac{1.808}{\sqrt{936.17}}\,,\ 0.257+2.05\cdot\frac{1.808}{\sqrt{936.17}}\right)\\&=(0.136\,,\ 0.378)\,.\end{aligned}$$

(c)

$$
\begin{aligned}
I_{y_0} &= y_{0\,obs} \pm t_{a/2}(n-2) \cdot s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{S_{xx}}} \\
&= (133.420 - 2.05 \cdot 1.808 \cdot 1.057\,,\ 133.420 + 2.05 \cdot 1.808 \cdot 1.057) \\
&= (129.50\,,\ 137.34)\,.
\end{aligned}
$$

(6) (a) Likelihoodfunktionen är

$$L(\theta) = \prod_{i=1}^{n} f_X(x_i;\theta) = \frac{1}{\theta^{2n}} \cdot x_1 \ldots x_n \cdot e^{-(x_1+\ldots+x_n)/\theta}$$

och logaritmerar vi så får vi

$$\ln L(\theta) = \sum_{i=1}^{n} \ln f_X(x_i;\theta) = -2n\ln\theta + \sum_{i=1}^{n} \ln x_i - \frac{1}{\theta}\sum_{i=1}^{n} x_i.$$

Deriverar vi detta, sätter derivatan $= 0$ och löser ut θ så fås

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{2n} = \frac{\overline{x}}{2}.$$

Kontroll ger även att andraderivatan är negativ, vilket garanterar att det verkligen är en max-punkt för likelihoodfunktion vi hittat.

(b) Vi har att

$$E[\hat{\theta}] = \frac{E[\bar{X}]}{2} = \frac{E[X]}{2},$$

där

$$E[X] = \int_0^\infty x f_X(x)\,dx = \int_0^\infty \frac{x^2}{\theta^2} e^{-x/\theta}\,dx = \theta \int_0^\infty y^2 e^{-y}\,dy = 2\theta$$

(t.e. via partiell integration). Alltså är

$$E[\hat{\theta}] = \theta,$$

dvs skattningen är väntevärdesriktig.